DOI:10.11918/202507019

可解释深度学习识别膜污染影响因子

陈君岩,周志伟,李 星,赵元添,李晓滢,王天阳

(北京工业大学 建筑工程学院,北京 100124)

摘 要:膜污染严重制约超滤(UF)技术在城市污水处理中的应用,成分复杂的二级出水有机物(EfOM)是其主要诱因,但关键膜污染组分尚缺乏有效的识别方法。为此,基于 TabPFN(Tabular Prior-Data Fitted Network)深度学习模型,结合 SHAP(SHapley Additive exPlanations)方法,构建关键膜污染因子识别框架。以改性陶瓷超滤膜(MCUM)2 h 运行周期内的水质与膜污染数据为基础,选取 UV_{254}、溶解性有机碳(DOC)、荧光区域积分值及荧光组分强度(F_{max})为输入,预测末端比通量、可逆阻力与不可逆阻力。结果显示,TabPFN 模型无需超参数调整即可实现高精度预测(末端比通量 $R^2=0.97$),整体性能优于传统机器学习模型。进一步引入后验集成(PHE)策略,预测性能提升至 $R^2=0.98$。SHAP 分析表明,色氨酸类、富里酸类和腐殖酸类物质分别在膜通量下降、不可逆阻力和可逆阻力的形成中具有主要影响。二维偏依赖图(PDPs)结果进一步揭示了这些组分之间显著的非线性协同作用,特别是富里酸类与芳香性物质在不可逆污染过程中的协同增强效应。本研究揭示了 EfOM 中主导膜污染的关键组分,为膜污染控制策略优化与污染机制理解提供了理论支撑。

关键词:膜污染;二级出水有机物;深度学习;可解释性分析;TabPFN 模型

中图分类号:X703.1 **文献标志码**:A **文章编号**:0367-6234(2026)06-0101-08

Interpretable deep learning for membrane fouling factor identification

CHEN Junyan, ZHOU Zhiwei, LI Xing, ZHAO Yuantian, LI Xiaoying, WANG Tianyang

(The College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: Membrane fouling seriously restricts the application of ultrafiltration (UF) technology in municipal wastewater treatment. Secondary effluent organic matter (EfOM), characterized by its complex composition, is considered the primary cause of membrane fouling. However, there is no method to effectively identify the key components of membrane fouling. To address this, this paper proposed a key membrane fouling factor identification framework based on the tabular prior-data fitted network (TabPFN) deep learning model by using SHapley Additive exPlanations (SHAP) method. Water quality and membrane fouling data were collected from a modified ceramic ultrafiltration membrane (MCUM) during a two-hour operation cycle. UV_{254}, dissolved organic carbon (DOC), fluorescence regional integration values, and fluorescence component intensities (F_{max}) were selected as input variables to predict normalized specific flux, reversible resistance, and irreversible resistance. The results show that the TabPFN model achieves high-precision prediction without hyperparameter tuning ($R^2=0.97$ for normalized specific flux), outperforming conventional machine learning models in terms of overall performance. Further improvement is obtained by introducing a post-hoc ensembling (PHE) strategy, increasing the prediction performance of R^2 to 0.98. SHAP analysis indicates that tryptophan-like, fulvic acid-like, and humic acid-like substances have predominant influences on the reduction of membrane flux and the formation of irreversible resistance and reversible resistance, respectively. PDPs results further reveal significant nonlinear synergistic effects among these components, especially the synergistic enhancement effect of the fulvic acid-like and aromatic substances in the irreversible fouling process. This study reveals the key components of secondary EfOM responsible for membrane fouling, providing theoretical support for membrane fouling control strategy optimization and the understanding of fouling mechanisms.

Keywords: membrane fouling; secondary effluent organic matter; deep learning; interpretable analysis; TabPFN model

超滤(UF)因其良好的截留性能,可有效去除水中的悬浮物、胶体、有机大分子及微生物,已广泛应用于污水处理领域[1]。但城市污水中的二级生化出水含有成分复杂的有机物(EfOM),易与膜材料发

收稿日期:2025-07-08;**录用日期**:2025-08-30;**网络首发日期**:2025-10-23
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20251023.0936.002
基金项目:国家自然科学基金(52370022)
作者简介:陈君岩(2001—),男,硕士研究生;周志伟(1985—),男,教授,博士生导师
通信作者:周志伟,zhouzw@bjut.edu.cn

生吸附或结合,导致膜污染、通量下降和运维成本上升[2]。EfOM 诱导的膜污染已成为制约超滤膜长期稳定运行的关键因素[3]。因此,识别 EfOM 中主导膜污染的关键组分,是深入理解膜污染机制与膜污染控制的基础。Zheng 等[4]通过 XAD 树脂将 EfOM 分离为 5 个组分,研究了各组分及其协同作用对膜污染的影响;Li 等[5]利用皮尔逊相关分析探讨了多种 EfOM 特征参数(如 UV_{254}、DOC、SUVA、荧光组分强度 F_{max} 和分子质量分布)与膜污染的关系。这些研究为理解 EfOM 诱导的膜污染提供了重要参考,但仍存在以下局限:一方面,实际水体成分复杂且动态变化[6],污染物间可能存在协同作用[7],导致配制水样难以全面反映真实的膜污染过程;另一方面,传统线性分析方法难以揭示多变量间潜在的非线性关联和高维交互机制[8],影响了关键污染因子的识别。近年来,机器学习在处理复杂、动态、非线性环境问题中展现出独特优势,已应用于膜污染预测以揭示水质参数与膜污染间的复杂关系[9-10]。然而,现有研究在关键污染因子的系统性识别与模型解释性方面仍存在不足。本研究基于膜滤出水中的 EfOM 特征参数与膜污染指标,构建了 TabPFN(Tabular Prior-Data Fitted Network)预测模型,实现对膜污染的精准预测。通过与传统机器学习方法对比,系统评估了 TabPFN 的性能优势。结合 SHAP(SHapley Additive exPlanations)方法与二维偏依赖图(PDPs)分析,进一步揭示了影响膜污染的关键水质因子,为膜污染主导因子的识别与污染控制策略优化提供理论依据。

1 模型与方法

1.1 数据集构建

本研究所用 Bi_2O_3/rGO/TiO_2(BRGT)改性陶瓷超滤膜(MCUM)基于课题组前期研究制备。具体而言,BRGT 三元异质结复合材料通过一步水热法合成[11],该材料兼具优异的吸附性能与光催化活性,能够协同促进水体中典型有机污染物的富集与降解。随后,采用真空抽滤结合热处理工艺,将 BRGT 负载于陶瓷膜表面,制备得到 MCUM[12]。实验以实际二级生化出水为处理对象,设置多种运行条件,包括不同光照模式(暗吸附、暗吸附饱和-紫外、紫外)、不同紫外光功率密度、不同 BRGT 负载量的 MCUM,不同剂量的过一硫酸盐(PMS)或过二硫酸盐(PDS)。在各运行条件下,收集 MCUM 处理后的渗透液,测定水质指标包括 UV_{254} 吸光度、可溶性有机碳(DOC)以及三维荧光光谱。通过扣除纯水空白消除瑞利和拉曼散射干扰,并将荧光强度统一归一化为拉曼单位(R. U.)[13]。随后,为捕捉 EfOM 中不同层级的荧光信息,同时采用区域荧光积分(FRI)方法与平行因子分析(PARAFAC)方法解析三维荧光数据,其中,FRI 侧重反映宏观区域强度变化,PARAFAC 则能更精细地识别独立荧光组分。在此基础上,三维荧光光谱被划分为 5 个特征区域:Ⅰ区(酪氨酸类蛋白)、Ⅱ区(色氨酸类蛋白)、Ⅲ区(富里酸类物质)、Ⅳ区(微生物代谢副产物类物质)和Ⅴ区(腐殖酸类物质),并采用 FRI 方法计算各特征区域的积分值。此外,结合 PARAFAC 方法进一步提取 3 种荧光组分,即微生物源类腐殖质(C1)、色氨酸类物质(C2)和陆源类腐殖质(C3),并记录各组分的最大荧光强度(F_{max})。最终,以 UV_{254}、DOC、5 个 FRI 值及 3 种荧光组分的 F_{max} 作为模型输入变量,膜运行 2 h 后的末端比通量、可逆阻力与不可逆阻力作为输出变量,构建 108 组样本数据用于模型训练与性能分析。

1.2 实验环境搭建

机器学习模型的训练与计算在 Windows 操作系统环境下进行,使用 Python 3.9 编程平台。硬件配置包括英特尔 i7-12650H 处理器及 NVIDIA GeForce RTX 4060(8 GB)独立显卡。深度学习框架采用 PyTorch 2.1.0,CUDA 版本为 11.8。所有变量在训练前均进行标准化处理,转换为均值为 0、方差为 1 的分布,消除量纲影响。

1.3 深度学习模型

TabPFN 是一种面向表格数据的元学习模型,融合了先验数据拟合机制与 Transformer 架构,其核心优势在于基于结构因果模型的合成数据生成策略,以及无需梯度更新即可完成预测的高效前向推理框架[14]。实验中,数据集按 7∶3 划分为训练集与测试集。TabPFN 模型在默认参数配置下,单次预测平均耗时约 1.8 s。为进一步提升模型性能,引入后验集成(PHE)策略。PHE 通过设定不同时间限制(30、120、300、900、1 800、3 600 s),在限定时间内多次验证不同参数配置的 TabPFN 模型,并采用贪婪集成算法对各子模型结果进行加权融合。

1.4 常规机器学习模型

选取 3 种常用机器学习算法,包括岭回归(Ridge)、径向基神经网络(RBF)和极端梯度提升树(XGBoost)。数据集按 7∶3 划分为训练集与测试集。为避免数据泄漏并提升模型泛化能力,针对不同算法设定合理的超参数搜索范围,结合三折交叉验证与遗传算法在训练集上进行超参数优化,最终在测试集上评估模型性能。

1.5　模型评价

为评估模型的预测性能，选用两种常用的回归评价指标决定系数（R^2）与均方根误差（E_{RMS}）。具体计算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$E_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

式中：y_i为第 i 个样本的实际值，$\hat{y}_i$为模型对第 i 个样本的预测值，$\bar{y}$为数据集中所有样本的平均值。R^2越接近 1，表明模型对数据的拟合程度越高，E_{RMS}越小，则表明预测误差越低。

1.6　特征评估

仅依赖机器学习模型的预测结果，难以深入解析输入特征与目标变量之间的潜在因果关系。SHAP 方法基于博弈论原理，能够为每个特征分配稳定且一致的贡献值，从而揭示其在模型预测中的作用机制[15]。相比传统特征重要性分析，SHAP 适用于多种复杂的非线性模型，具备全局和局部解释能力，可同时识别整体上的主要驱动因子和单一样本的关键影响路径[16-17]，为量化水质参数对膜污染的贡献程度提供了有效手段。平均 SHAP 绝对值可反映各特征对模型输出的平均贡献水平，数值越高，表明该特征在预测结果中的影响越显著[18]。因此，SHAP 值较高的水质参数可视为影响末端比通量、可逆阻力与不可逆阻力的关键因素，即膜污染过程中的主导因子。此外，PDPs 通过展示两个特征变量对预测结果的联合影响[19]，辅助揭示特征间的非线性关系和交互效应，补充 SHAP 的贡献度分析。

2　结果与讨论

2.1　TabPFN 模型与传统机器学习模型性能对比

在水质参数与末端比通量预测任务中（表 1），RBF 神经网络在默认超参数下预测能力有限（测试集 $R^2 = 0.71$）。引入遗传算法进行超参数优化后，模型预测性能显著提升，测试集 R^2 提高至 0.95，E_{RMS}降至 0.029。XGBoost 模型虽在训练集上拟合度极高（$R^2 = 1.00$），但测试集表现明显下降（$R^2 = 0.93$），存在过拟合现象。通过遗传算法优化，XGBoost 泛化能力改善，测试集 R^2提升至 0.94，E_{RMS}降至 0.030，表明超参数优化可有效增强非线性模型的预测性能与泛化能力。相比之下，经过遗传算法优化后的线性回归模型（Ridge 回归）虽具备一定的预测能力（$R^2 = 0.88$，$E_{RMS} = 0.044$），但仍难以有效拟合输入变量与目标值之间的非线性关系，整体预测效果低于 RBF 与 XGBoost 模型。值得注意的是，TabPFN 模型在未进行任何超参数调整的情况下，仅通过一次前向传播即实现了最优的末端比通量预测效果（$R^2 = 0.97$，$E_{RMS} = 0.021$）。

表 1　训练集与测试集上的末端比通量预测性能评估

Tab. 1　Prediction performance evaluation of normalized specific flux on training and testing sets

模型	训练集				测试集			
	默认参数		遗传算法优化		默认参数		遗传算法优化	
	R^2	E_{RMS}	R^2	E_{RMS}	R^2	E_{RMS}	R^2	E_{RMS}
Ridge	0.90	0.040	0.91	0.039	0.87	0.045	0.88	0.044
RBF	0.84	0.051	0.96	0.025	0.71	0.067	0.95	0.029
XGBoost	1.00	0	1.00	0.006	0.93	0.034	0.94	0.030
TabPFN	0.99	0.015			0.97	0.021		

进一步分析可逆与不可逆阻力的预测结果（见表 2、3），Ridge 与 RBF 模型在默认超参数条件下预测能力有限（测试集 R^2低于 0.75）。引入遗传算法进行超参数优化后，两者性能均有所提升，尤其是在不可逆阻力预测任务中，RBF 模型的测试集 R^2提升至 0.82。XGBoost 在两类阻力预测中同样存在一定的过拟合风险，且在不可逆阻力预测中，超参数优化后测试集性能出现小幅下降（R^2由 0.87 降至 0.85）。说明传统机器学习模型的超参数优化并不总能提升测试集预测性能，其效果受限于数据集复杂性和优化策略。相比之下，TabPFN 无需调参，在两类阻力预测任务中均表现较高的测试集 R^2，分别为 0.95（可逆阻力）和 0.86（不可逆阻力），体现出更高效且鲁棒的预测能力。

表 2 训练集与测试集上的可逆阻力预测性能评估

Tab. 2 Prediction performance evaluation of reversible resistance on training and testing sets

模型	训练集				测试集			
	默认参数		遗传算法优化		默认参数		遗传算法优化	
	R^2	$E_{RMS}/10^8$	R^2	$E_{RMS}/10^{10}$	R^2	$E_{RMS}/10^{11}$	R^2	$E_{RMS}/10^{11}$
Ridge	0.74	6 350	0.75	62.2	0.72	6.08	0.74	5.83
RBF	0.58	7 980	0.80	55.0	0.53	7.82	0.77	5.49
XGBoost	1.00	8.93	1.00	1.07	0.76	5.58	0.91	3.52
TabPFN	0.98	1 810			0.95	2.58		

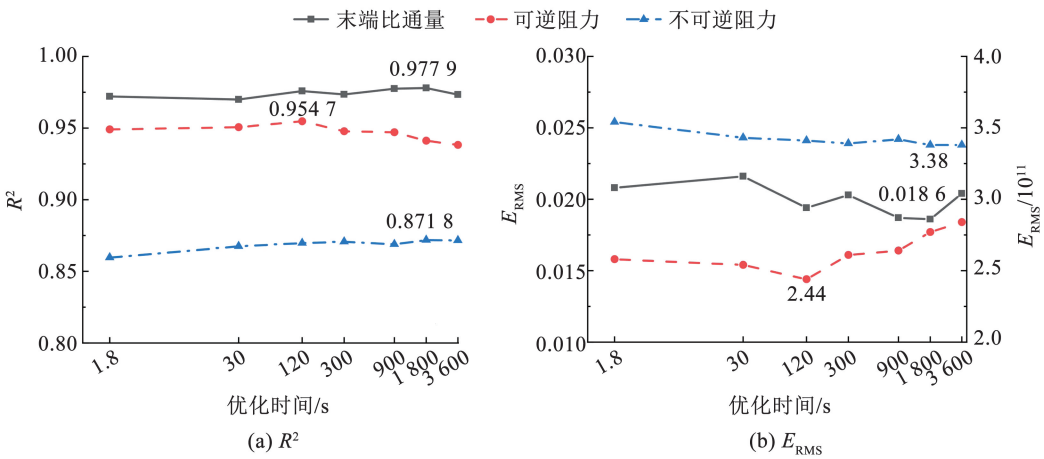
表 3 训练集与测试集上的不可逆阻力预测性能评估

Tab. 3 Prediction performance evaluation of irreversible resistance on training and testing sets

模型	训练集				测试集			
	默认参数		遗传算法优化		默认参数		遗传算法优化	
	R^2	$E_{RMS}/10^9$	R^2	$E_{RMS}/10^8$	R^2	$E_{RMS}/10^{11}$	R^2	$E_{RMS}/10^{11}$
Ridge	0.81	540	0.83	5 200	0.74	4.85	0.75	4.72
RBF	0.76	616	0.94	3 050	0.57	6.17	0.82	4.03
XGBoost	1.00	1.07	1.00	8.59	0.87	3.44	0.85	3.70
TabPFN	0.96	240			0.86	3.54		

2.2 TabPFN(PHE)模型的性能提升分析

如图 1 所示,TabPFN(PHE)的预测性能显著依赖集成时间,合理选择集成时间有助于在保证预测精度的同时控制计算成本。末端比通量和不可逆阻力任务在 1 800 s 达到最佳表现,分别为 $R^2=0.98$, $E_{RMS}=0.018\ 6$ 和 $R^2=0.87$, $E_{RMS}=3.38\times10^{11}$;可逆阻力任务则在 120 s 达到最优($R^2=0.95$, $E_{RMS}=2.44\times10^{11}$)。各任务均呈现出预测效果随集成时间增加而提升,随后略有下降的趋势,表明适度集成能够增强模型泛化能力,但过长集成时间可能引入表现较差的 TabPFN 配置,导致集成效果下降。

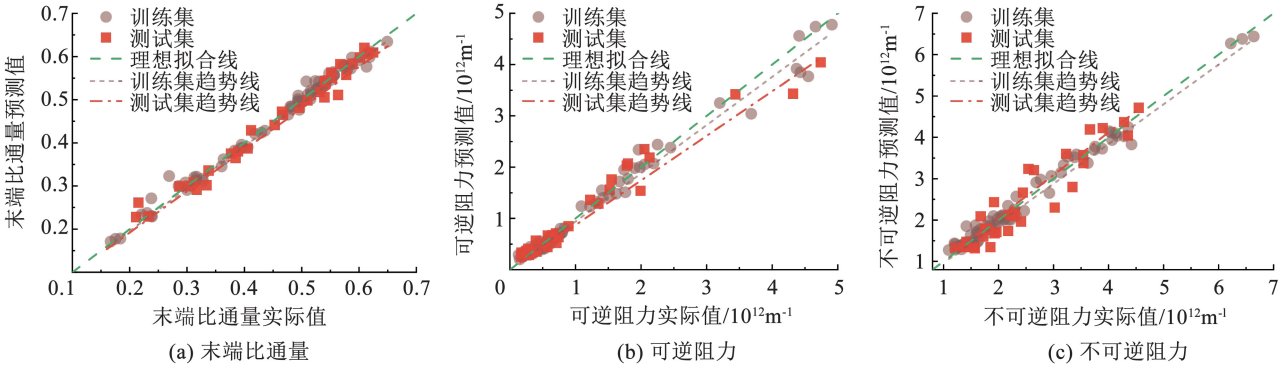


(a) R^2 (b) E_{RMS}

图 1 TabPFN(PHE)模型在不同优化时间下测试集预测性能对比

Fig. 1 Prediction performance comparison of TabPFN (PHE) model on testing set under different optimization durations

图 2 展示了 TabPFN(PHE)在各任务最优集成时间下对目标变量的预测效果。末端比通量与不可逆阻力任务中,训练集与测试集趋势线斜率分别为 0.98、0.96 和 0.96、1.03,预测值主要分布在理想拟合线附近,整体斜率接近 1,表明模型预测结果与真实值具有良好的线性一致性。可逆阻力任务中,训练集、测试集斜率为 0.94、0.87,测试集预测结果相对偏低,主要受可逆阻力大于 $3\times10^{12}\ m^{-1}$ 区间预测误差的影响。



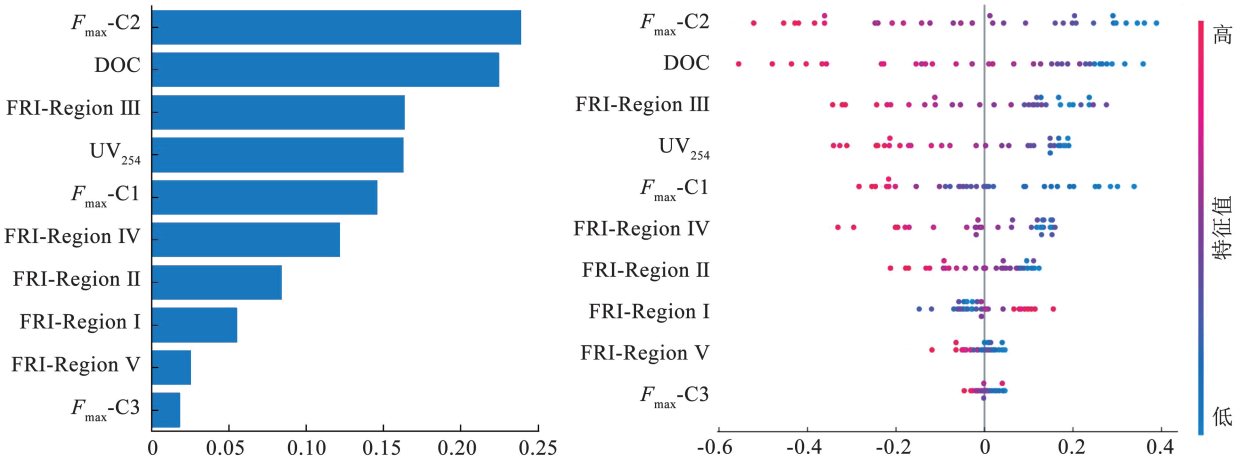
(a) 末端比通量　(b) 可逆阻力　(c) 不可逆阻力

图 2　最优 TabPFN(PHE)模型的预测性能

Fig. 2　Prediction performance of optimal TabPFN (PHE) model

2.3　最优 TabPFN(PHE)模型的解释性分析

基于 2.2 节确定的各任务最优集成时间，进一步采用 SHAP 方法对相应的 TabPFN(PHE)模型进行了特征重要性分析，以识别关键膜污染因子。如图 3～5 所示，DOC 在 3 类污染预测任务中的 SHAP 值分别为 0.22、0.40 和 0.57，表明其在膜污染形成过程中具有重要影响。DOC 作为水样中有机负荷的综合性指标，可能包括大分子(>10 000 u)生物聚合物、中等分子(1 000～10 000 u)的腐殖质物质以及低分子(<1 000 u)有机酸等有机组分[20]。其中，生物聚合物(如蛋白质和多糖)易在膜表面形成凝胶层，覆盖膜孔，引发可逆性与不可逆性膜污染；中等分子质量的疏水性腐殖质类物质则倾向于在膜表面积聚，形成致密滤饼层，导致可逆污染；而低分子有机酸由于尺寸较小，易渗透进入膜孔内部，通过吸附与堵塞作用引起不可逆污染[4,21]。因此，较高的 DOC 反映了进入膜表面与孔隙的有机物总量增加，这个过程加剧了膜污染，最终导致比通量下降，可逆阻力及不可逆阻力的上升。



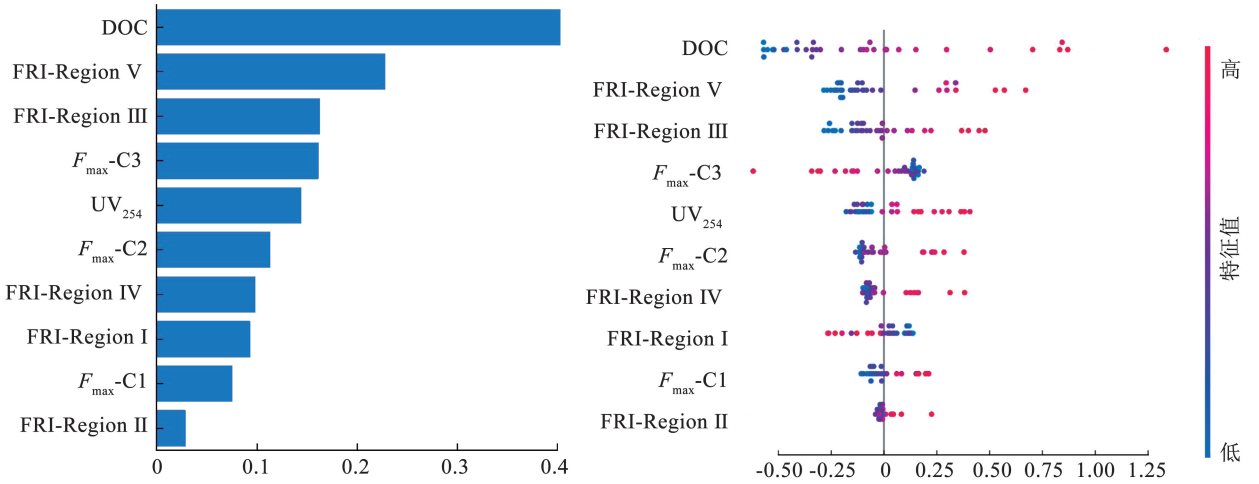
(a) 平均SHAP值　(b) SHAP值

图 3　末端比通量的 SHAP 特征重要性分析

Fig. 3　SHAP feature importance analysis of normalized specific flux

针对具体膜污染指标，SHAP 分析揭示了不同 EfOM 组分在污染形成过程中的差异化贡献。在末端比通量的预测中(图 3)，F_{max}-C2(SHAP 值为 0.24)主导了通量的下降。这主要是由于膜表面的含氧官能团易与蛋白质分子发生强相互作用[22]，从而促使 EfOM 中的色氨酸蛋白样物质更容易与 MCUM 表面负载的还原氧化石墨烯(rGO)吸附或结合，进而加剧膜表面污染。Fan 等[23]的研究同样表明 C2 组分与膜污染指标之间的皮尔逊相关系数高达 0.99，进一步验证了其在通量下降中的主导作用。在可逆阻力的形成中，SHAP 分析(图 4)显示，腐殖酸类物质(FRI-Region V，SHAP 值为 0.23)和陆源类腐殖质组分(F_{max}-C3，SHAP 值为 0.16)具有显著影响。这主要是由于分子质量大于 50 000 u 的腐殖酸类物质易在膜表面吸附并形成疏松的滤饼层，在过滤压力作用下产生弱结合的污染层[24]。这类污染通常以可逆形式存在，可通过反冲洗或温和的化学清洗方式去除。在不可逆阻力预测中

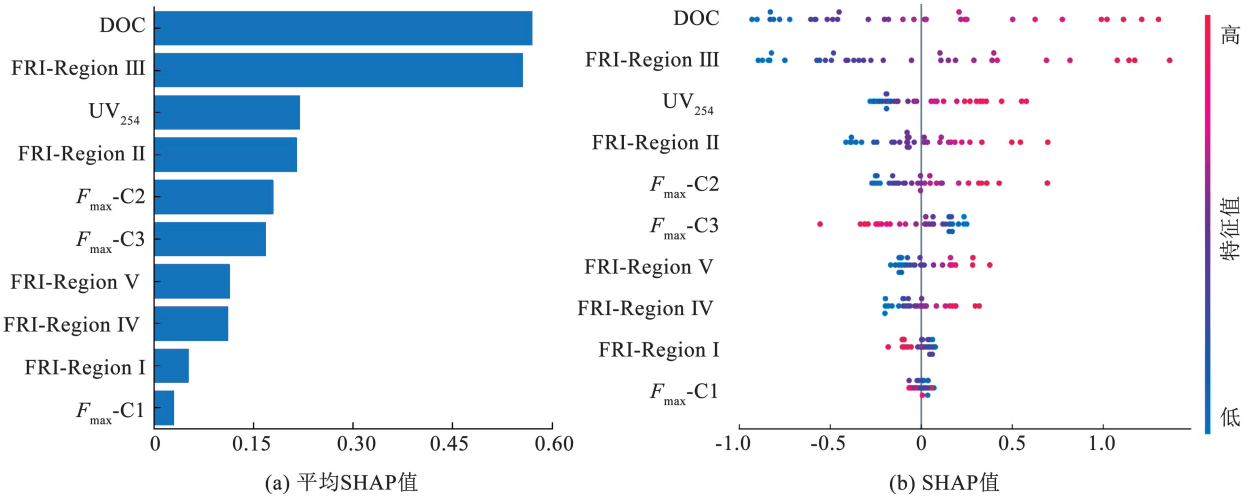
(图 5),富里酸类物质(FRI-Region III)的 SHAP 值高达 0.56,仅次于 DOC,表明其在不可逆污染形成中具有关键作用。这主要归因于 EfOM 中的中低分子质量(1 000 ~ 3 000 u)富里酸物质在膜过滤过程中易穿透滤饼层并进入膜孔,造成孔道堵塞,最终形成不可逆污染[25]。



(a) 平均SHAP值 (b) SHAP值

图 4 可逆阻力的 SHAP 特征重要性分析

Fig. 4 SHAP feature importance analysis of reversible resistance



(a) 平均SHAP值 (b) SHAP值

图 5 不可逆阻力的 SHAP 特征重要性分析

Fig. 5 SHAP feature importance analysis of irreversible resistance

为进一步探究关键膜污染因子之间的协同作用,基于 SHAP 特征重要性分析结果,选取了贡献度较高的特征变量,构建了 PDPs 以可视化不同变量组合对膜污染指标的综合影响。在末端比通预测中(图 6(a) ~6(c)),3 组变量组合的等高线形态和颜色分布较为一致,且当两种指标数值最高时,膜通量显著下降至最低值(末端比通量约 0.30),均表现出明显的协同负效应。对于可逆阻力(图 6(d) ~ 6(f)),变量间的交互模式与末端比通量相似,但整体协同效应逐渐减弱,颜色条数值范围也相应减小。DOC 与腐殖酸类物质的组合显示出最强的交互作用,对可逆阻力的影响最强,预测值波动最大(0.8×10^{12} ~4.8×10^{12} m^{-1}),说明他们之间协同效应显著。相比之下,DOC 与富里酸类以及腐殖酸与富里酸类的组合影响较弱,预测值变化范围较小。这表明当 DOC 质量浓度升高,尤其腐殖酸类含量较高时,可逆污染更易形成,从而显著增加可逆阻力。在不可逆阻力分析中,不同污染物组合表现出差异化的交互特性。而在图 6(g)中,不可逆阻力峰值($>9.0\times10^{12}$ m^{-1})集中于 DOC 和富里酸类物质均高的区域,显示其强协同对不可逆污染的显著影响。图 6(h)中,等高线沿 DOC 轴分布,表明该指标主要由 DOC 单因子主导,UV_{254} 的协同效应较弱。值得注意的是,尽管图 6(h)中包含最重要因子 DOC,其预测值范围(1.8×10^{12} ~ 6.0×10^{12} m^{-1})小于图 6(i)中富里酸类物质与 UV_{254} 的组合(2.4×10^{12} ~ 7.2×10^{12} m^{-1})。这表明在不可逆污染的形成过程中,富里酸类物质与芳香性物质这两类次重要因子的强协同作用,可能对污染程度产生的影响超过 DOC 等主导因子的单独或弱协同效应。

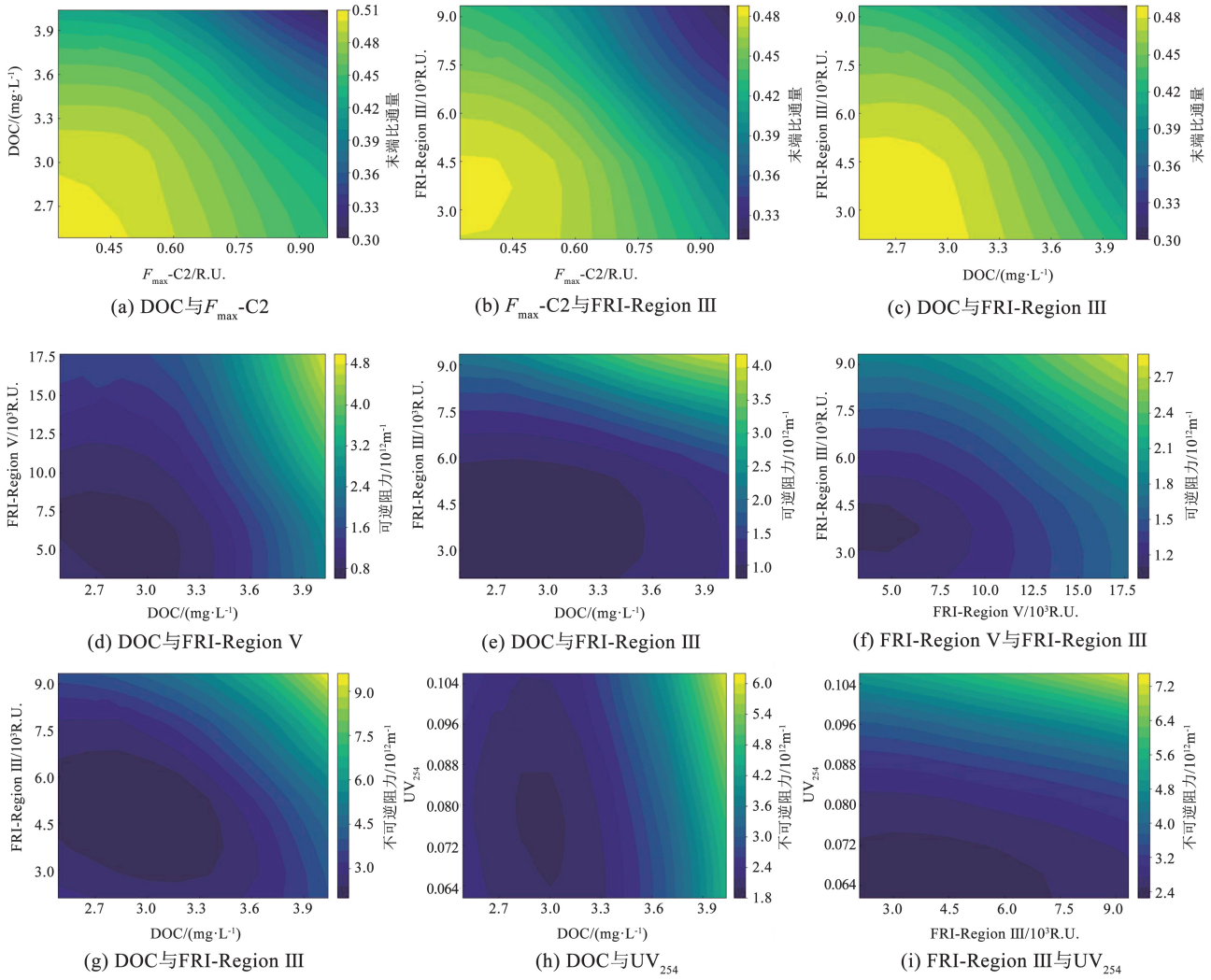


图 6　基于 SHAP 特征重要性的关键变量对膜污染指标(末端比通量、可逆、不可逆阻力)的二维偏依赖分析

Fig. 6　Two-dimensional partial dependence analysis of key variables on membrane fouling metrics (normalized specific flux, reversible, irreversible resistance) based on SHAP feature importance

综上，SHAP 值分析揭示了色氨酸类、富里酸类和腐殖酸类物质为膜污染的主要驱动因子。其中，色氨酸类物质显著驱动末端比通量下降，富里酸类物质对不可逆阻力的形成具有关键影响，而腐殖酸类物质易在膜表面形成滤饼层导致可逆阻力上升。PDPs 分析进一步揭示了这些组分间存在显著的非线性协同效应，特别是在高有机负荷条件下，色氨酸类与腐殖酸类物质对膜通量和可逆阻力的协同加剧作用明显；富里酸类物质与芳香性物质的强协同效应则对不可逆阻力的加重起重要作用。基于关键膜污染因子的识别，应优先提升 MCUM 表面负载 BRGT 对有机污染物的富集及降解能力，尤其针对主要膜污染驱动因子，同时强化 BRGT 复合材料对易引发膜污染的 EfOM 组分的膜表面原位降解能力，从而有效减少污染物积累和膜污染。此外，不同氧化剂对不同类型污染物具有特异性降解能力[26]，MCUM 可结合特定氧化剂，针对高污染倾向的污染物实现进一步去除，从而有效缓解膜污染。

3　结　论

1) 在末端比通量、可逆阻力与不可逆阻力的预测任务中，传统模型(Ridge、RBF、XGBoost)即使利用遗传算法进行超参数优化，整体预测效果仍然有限，且部分模型存在过拟合风险。相比之下，TabPFN 模型无需超参数调整即可实现优于传统模型的预测效果，体现出其在非线性关系建模与泛化能力方面的优势。

2) 引入 TabPFN(PHE)模型后，不同预测任务的性能进一步提升。末端比通量与不可逆阻力在 1 800 s集成时间下达到最优效果，可逆阻力的最优效果出现在 120 s 集成时间。适度集成可增强模型鲁棒性，过度集成则可能导致预测性能下降。

3) SHAP 分析识别了二级出水有机物(EfOM)中导致膜污染的关键组分，主要包括色氨酸类物质、

富里酸类物质和腐殖酸类物质,并分析了其在膜通量下降以及可逆和不可逆阻力形成过程中的不同贡献机制。

4)PDPs 分析揭示,关键有机组分在较高有机负荷(如 DOC 质量浓度升高)条件下呈现出显著的非线性协同效应。具体而言,色氨酸类物质主要促使膜通量下降,腐殖酸类物质显著增加可逆阻力,富里酸类和芳香性物质则更易加剧不可逆阻力。这些结果加深了对膜污染机制的理解,为后续防控策略提供了科学依据。

参考文献

[1] QU Jiuhui, DAI Xiaohu, HU Hongying, et al. Emerging trends and prospects for municipal wastewater management in China[J]. ACS ES&T Engineering, 2022, 2(3): 323. DOI: 10.1021/acsestengg.1c00345

[2] SONG Zilong, WANG Zhenbei, MA Jun, et al. Molecular levels unveil the membrane fouling mitigation mechanism of a superpotent N-rGO catalytic ozonation membrane: interfacial catalytic reaction pathway and induced EfOM transformation reactions[J]. Applied Catalysis B: Environmental, 2022, 319: 121943. DOI: 10.1016/j.apcatb.2022.121943

[3] ZHENG Xing, ERNST M, JEKEL M. Identification and quantification of major organic foulants in treated domestic wastewater affecting filterability in dead-end ultrafiltration[J]. Water Research, 2009, 43(1): 238. DOI: 10.1016/j.watres.2008.10.011

[4] ZHENG Xing, KHAN M T, CROUE J P. Contribution of effluent organic matter (EfOM) to ultrafiltration (UF) membrane fouling: isolation, characterization, and fouling effect of EfOM fractions[J]. Water Research, 2014, 65: 414. DOI: 10.1016/j.watres.2014.07.039

[5] LI Peijie, XU Daliang, GAO Yunfei, et al. Nano-confined catalysis with Co nanoparticles-encapsulated carbon nanotubes for enhanced peroxymonosulfate oxidation in secondary effluent treatment: water quality improvement and membrane fouling alleviation[J]. Water Research, 2024, 266: 122357. DOI: 10.1016/j.watres.2024.122357

[6] O'CONNOR M, HELAL S R, LATCH D E, et al. Quantifying photo-production of triplet excited states and singlet oxygen from effluent organic matter[J]. Water Research, 2019, 156: 23. DOI: 10.1016/j.watres.2019.03.002

[7] XIAO Feng, XIAO Peipei, ZHANG Weijun, et al. Identification of key factors affecting the organic fouling on low-pressure ultrafiltration membranes[J]. Journal of Membrane Science, 2013, 447: 144. DOI: 10.1016/j.memsci.2013.07.040

[8] HOWELL D C. Statistical methods for psychology[M]. [S.l.]: PWS-Kent Publishing Co, 1992

[9] ZHENG Wenjing, CHEN Yan, XU Xiaohu, et al. Research on the factors influencing nanofiltration membrane fouling and the prediction of membrane fouling[J]. Journal of Water Process Engineering, 2024, 59: 104876. DOI: 10.1016/j.jwpe.2024.104876

[10] WANG Lu, LI Zonghao, FAN Jianhua, et al. The intelligent prediction of membrane fouling during membrane filtration by mathematical models and artificial intelligence models[J]. Chemosphere, 2024, 349: 141031. DOI: 10.1016/j.chemosphere.2023.141031

[11] YU Rui, YANG Yanling, ZHOU Zhiwei, et al. Facile synthesis of ternary heterojunction Bi_2O_3/reduced graphene oxide/TiO_2 composite with boosted visible-light photocatalytic activity[J]. Separation and Purification Technology, 2022, 299: 121712. DOI: 10.1016/j.seppur.2022.121712

[12] ZHOU Zhiwei, CHEN Junyan, YU Haikuan, et al. Bi_2O_3/rGO/TiO_2 catalytic membrane for advanced treatment of secondary effluent: insights into EfOM degradation and membrane fouling alleviation[J]. Journal of Water Process Engineering, 2025, 77: 108933. DOI: 10.1016/j.jwpe.2025.108399

[13] LAWAETZ A J, STEDMON C A. Fluorescence intensity calibration using the Raman scatter peak of water[J]. Applied Spectroscopy, 2009, 63(8): 936. DOI: 10.1366/000370209788964548

[14] HOLLMANN N, MULLER S, PURUCKER L, et al. Accurate predictions on small data with a tabular foundation model[J]. Nature, 2025, 637(8045): 319. DOI: 10.1038/s41586-024-08328-6

[15] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[J]. Advances in Neural Information Processing Systems, 2017(2): 30. DOI:10.48550/arXiv.1705.07874

[16] PANG Hongjiao, BEN Yawen, CAO Yong, et al. Time series-based machine learning for forecasting multivariate water quality in full-scale drinking water treatment with various reagent dosages[J]. Water Research, 2025, 268(Pt B): 122777. DOI: 10.1016/j.watres.2024.122777

[17] HE Song, DU Wenli, PENG Xin. Uncertainty-aware optimization of zeolite particle size: utilizing quantile regression and SHAP analysis[J]. Chemical Engineering Journal, 2025, 512: 162085. DOI: 10.1016/j.cej.2025.162085

[18] 徐达梁, 徐杭镔, 靳心瑶, 等. 基于机器学习的纳滤膜预测筛选模型构建与评估[J]. 哈尔滨工业大学学报, 2024, 56(6): 8
XU Daliang, XU Hangbin, JIN Xinyao, et al. Construction and evaluation of prediction model for nanofiltration membranes based on machine learning[J]. Journal of Harbin Institute of Technology, 2024, 56(6): 8. DOI: 10.11918/202402004

[19] DENG Ligang, LIU Kai, FAN Yifan, et al. Interpretable machine learning models reveal the partnership of microplastics and perfluoroalkyl substances in sediments at a century scale[J]. Journal of Hazardous Materials, 2025, 486: 137018. DOI: 10.1016/j.jhazmat.2024.137018

[20] HU Jingyi, SHANG Ran, HEIJMAN B, et al. Influence of activated carbon preloading by EfOM fractions from treated wastewater on adsorption of pharmaceutically active compounds[J]. Chemosphere, 2016, 150: 49. DOI: 10.1016/j.chemosphere.2016.01.121

[21] GUO Jin, LIU Hong, LIU Jianhong, et al. Ultrafiltration performance of EfOM and NOM under different MWCO membranes: comparison with fluorescence spectroscopy and gel filtration chromatography[J]. Desalination, 2014, 344: 129. DOI: 10.1016/j.desal.2014.03.006

[22] HAN Yi, JIANG Yangqiu, GAO Chao. High-Flux graphene oxide nanofiltration membrane intercalated by carbon nanotubes[J]. ACS Applied Materials & Interfaces, 2015, 7(15): 8147. DOI: 10.1021/acsami.5b00986

[23] FAN Qingshui, CHENG Xiaoxiang, ZHU Xuewu, et al. Secondary wastewater treatment using peroxymonosulfate activated by a carbon nanofiber supported Co_3O_4(Co_3O_4@CNF) catalyst combined with ultrafiltration[J]. Separation and Purification Technology, 2022, 287: 120579. DOI:10.1016/j.seppur.2022.120579

[24] 张晓飞. 不同级分的腐殖酸混凝预处理控制及对膜污染的影响研究[D]. 哈尔滨: 哈尔滨工业大学, 2016
ZHANG Xiaofei. Study on coagulation pretreatment control of different humic acid fraction and the influence of membrane fouling process[D]. Harbin: Harbin Institute of Technology, 2016

[25] WU Wenjun, GUO Jin, WANG Wenhui, et al. Nanoconfined catalytic membranes for EfOM fouling mitigation: an intelligent "pore-centric" cleaning strategy[J]. Journal of Membrane Science, 2025, 715: 123444. DOI: 10.1016/j.memsci.2024.123444

[26] LI Mo, WEN Qinxue, ZHANG Yongming, et al. New insights into the transformation of effluent organic matter during Fe(Ⅱ)-assisted advanced oxidation processes: parallel factor analysis coupled with self-organizing maps[J]. Water Research, 2022, 221: 118789. DOI: 10.1016/j.watres.2020.116299

(编辑 刘 彤)